

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
28 February 2002 (28.02.2002)

PCT

(10) International Publication Number  
WO 02/17190 A1

(51) International Patent Classification: G06F 17/60

(21) International Application Number: PCT/US01/25956

(22) International Filing Date: 20 August 2001 (20.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
09/643,643 22 August 2000 (22.08.2000) US(71) Applicant: VARRO TECHNOLOGIES, INC. [US/US];  
325 Cherry Street, 1st floor, Philadelphia, PA 19106 (US).(72) Inventors: HEDGEPTH, Chester, III; 2300 Walnut  
Street, Apt. 727, Philadelphia, PA 19103 (US). SHIN,  
David, Sullivan; 318 Lincoln Woods, Lafayette Hill, PA  
19444 (US). VENKATESAN, Jay, Raman; 514 South  
22nd Street, Philadelphia, PA 19146 (US).(74) Agent: GOLUB, Daniel, H.; 1701 Market Street,  
Philadelphia, PA 19103 (US).(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MY, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,  
SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SI, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,  
TG).

## Declarations under Rule 4.17:

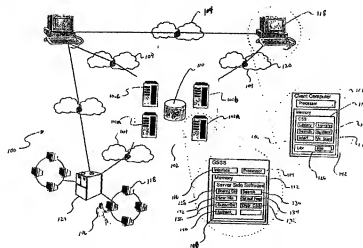
- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for all designations
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations

## Published:

- with international search report

[Continued on next page]

(54) Title: METHOD AND SYSTEM FOR SHARING BIOLOGICAL INFORMATION



(57) Abstract: The present invention is a system and method for sharing biological information such as genetic sequences. The system (100) is based on providing a host computer (102) which maintains a database of biological information available for sharing by members (106) of a biological information sharing network. The host computer (102) maintains a database of information available for sharing. Sharing is accomplished by transferring biological information from one member's computer (118) to another's. Software is preferably provided to the members (106) to assist in the formatting of biological information to be shared. The software may also have a virus scanning capability to prevent infected information files from being shared. Provisions may also be supplied for encrypting information being transferred from one member to another, or for using digital rights management technology for limiting the ability of the receiving member to re-share the biological information.



- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## METHOD AND SYSTEM FOR SHARING BIOLOGICAL INFORMATION

### Field of the Invention

The present invention pertains to the sharing of information related to biological research, and more particularly, to the sharing of genetic sequence information accumulated on individual computers through client computer to client computer file transfers.

### Background of the Invention

Research into genetic structures has provided a large mass of biological information, including but not limited to genetic sequence information, protein/peptide sequences, antibody data, cell line, ribozyme sequences, anti-sense oligo sequences, transgenic data, chemical structure/database information, and molecular beacon data.

Much of the biological information is characterized by a sequence of values defining the characteristics of the biological information. Deoxyribonucleic acid (DNA) is a molecule that carries genetic information, encoded in a linear sequence of nucleotide bases. There are four such bases, generally abbreviated A, C, T and G, so that DNA is generally represented as a long sequence drawn on a four letter alphabet. Acting in concert with the environment, these DNA molecules (called the genome or genotype) determine the structure, function and behavior (called the phenotype) of all living things.

Knowledge of these sequences is useful for many purposes, including reducing the expense associated with the testing of new drugs, and efforts to understand diseases

such as cancer and AIDS. This mass of knowledge, however, is distributed among many researchers and publications.

Several public databases, such as the dbGSS database provided by the National Center for Biotechnology Information, exist for publishing information related to  
5 genomics. These databases require an extensive review of biological information before the biological information may be posted to the database, resulting in delays before the biological information becomes available to other researchers. Also, the database host must be able to store all of the information submitted by the submitters of the biological information.

10 Other databases are run by private organizations, with the purpose of selling biological information for which the private organization has acquired proprietary rights. These private databases also require large storage capacities to enable the host computer to store all of the biological information referenced in the database.

Public and private databases are consistent in that each database includes only a  
15 limited amount of biological information, necessitated by the limited storage capacity of the database operator, the requirements for submitting information for the database, and proprietary interests in maintaining a database for information owned by the database host.

Recently, as a technique for sharing music files, programs have been developed  
20 which allow a host database to direct searchers to member hard drives to copy sought after music files. As such, the actual transfer of music files occurs between individual members of the network, commonly referred to as peer-to-peer sharing. The host database

includes a description of material available on individual member's hard drives, and the address of the member (which can be a floating IP address), such that other members may search for specific music by querying the host database for storage locations of music meeting parameters identified by the searching member.

5           As an evolutionary step beyond the use of a host database for sharing music files, software has been developed which allows full peer-to-peer sharing of information stored on individual member hard-drives, without requiring the intermediate step of searching a hosted database of characteristics identify available material to identify the locations of the material. The peer-to-peer sharing software allows an individual member to generate  
10 a query which is passed to the computer of each member in the peer-to-peer network, where the query is compared to material shared on the member computers.

The use of sharing technology has several disadvantages. The sharing technologies create security risks for both the requesting party and the sharing party in a transaction. The requesting member is able to place data on the sharing member's hard  
15 drive during the transfer, resulting in the potential for computer viruses to be distributed through the sharing network. Such viruses can also be transferred from the sharing member's computer to the requesting member's computer.

The existing sharing technology is also limited in that a requesting member does not receive any quantification or qualification of the accuracy of the information shared.  
20 Whereas major biological information databases screen submissions for accuracy, information shared by members in a peer-to-peer sharing network, either with or without a centralized database, becomes available for sharing without undergoing peer review. This

is both a benefit and a disadvantage, in that the reduced review requirement leads to less effort being required to share the data, while at the same time allowing inaccurate data to become published.

#### Summary of the Invention

5           The present invention is a system and method for sharing biological information over a network. The system includes members who are communicably connected over the network to a network host. At least one of the members has biological information stored on a computer that they are willing to share over the network. The network host has a database, which contains criteria identifying the biological information stored on  
10 member's computers. Associated with the criteria identifying the biological information stored on member's computers are references which identify where the information is stored. The network host provides a search tool for the members, so that the members may search the biological information shared by other members to find biological information of interest. Once a searching member has identified biological information of  
15 interest, a copy of the information may be transferred from the computer on which the information is stored onto the computer of the searching member.

          The members are provided with software for installation on their computers. The software assists in formatting biological information which individual members are willing to share, as well as reports to the network host what biological information has  
20 been shared by each member.

          The software further comprises features for enhancing the security of individual members associated with the network. The software in one embodiment includes a virus scanner which determines whether biological information made available for sharing is

infected with a computer virus, and if infected reports the infection to the network host.

In another embodiment, the software encrypts the biological information being transferred so that it may not be intercepted during transmission. In an alternate embodiment, digital rights management technologies are used to limit the use of the biological information by the searching member who receives the information.

The invention is also embodied in a method for exchanging biological information. The method is based on providing at least one host computer, and subscribing several members to form a biological information exchange network. Each member has a network device capable of storing biological information, and communicating with the host computer. A database is provided on the host computer. The database stores information describing biological information made available for sharing on the individual member's network devices and information regarding where shared biologic material is stored.

The host computer receives search parameters from a member seeking biological information, and searches the criteria provided by the subscribing members describing biological information which is available for sharing. In one embodiment, the host computer also determines whether biological information which meets the search criteria is stored on a member's network device that is communicably connected to the host computer, and generates a list of biological information available for sharing for the searching member, wherein the list contains only references to biological information on network devices that are communicably connected to the host computer. In this embodiment, only member-to-member transfers of copies of biological information are enabled.

In another embodiment, the one host computer generates a list of all biological information which is shared, and allows delayed transfer of a copy of the biological information when both searching and sharing members are communicably connected to the network host. Alternately, the host computer receives a copy of the biological  
5 information when the sharing member is communicably connected to the host computer, and transfers the copy to the searching member when the searching member is communicably connected with the host computer.

The present invention is also embodied in a business which provides a biological information exchange. The biological information exchange charges a fee for  
10 membership in the biological information exchange. The fee may be a flat fee based on membership in the exchange network, or based on the number of files transferred by an individual searching member.

#### Brief Description of the Drawings

Figure 1 shows a system diagram of a system embodying the present invention.

15 Figure 2 shows a database structure as used with the present embodiment.

Figure 2A shows an alternative database structure consistent with the present invention.

Figure 3 shows an illustrative embodiment of a client side software member interface, showing a search entry display.

20 Figure 4 shows an illustrative embodiment of a client side software member interface showing a file transfer display.



Figure 5 shows a process flowchart for providing a genetic information sharing system.

#### Detailed Description of the Invention

Referring to Figure 1, wherein like numerals refer to like elements, there is shown a diagram of a genetic Biological Information sharing system 100 according to the present invention, implemented to show genetic sequence information being exchanged. The genetic sequence sharing server 102 (hereafter "GSSS") is connected to a network 104 such as the Internet or a local intranet via a communications connection. The GSSS 102 of the embodiment illustrated is connected to the Internet. Such connections are commonly known, and may include dial-up connections, cable or DSL interfaces, or a dedicated interface such as a T1 connection. The GSSS 102 acts as a repository or database of references to members 106 who have indicated a willingness to share genetic information, and also hosts server side software 108 (hereafter "SSS") which controls upkeep and access to the database 110.

The GSSS 102 is generally a computer which includes memory 112, a processor 114, and an interface 116 to a network. In the presently illustrated embodiment, the network is the Internet, allowing researchers to become members of the system without requiring significant investment in communications infrastructure. Although the GSSS 102 is described in the illustrated embodiment as a single server, a cluster of servers 102a, 102b, ... 102n may be used to provide adequate system performance if the number of members 106 so requires. In such a configuration, each server is a copy of each other, with one server being designated as the master server.

The memory of the GSSS 102 stores sharing system software 108, described below. A database 110 of information regarding available genetic sequences is also connected to the GSSS 102, either resident in memory or housed on a separate machine. The database 110 may use a relational database for storing information related to individual shared genetic sequences, as described below. The database 110 may be a central database, with multiple sharing servers 102a, 102b... 102n able to access the database 110, or it may include mirrored copies of a database, with a copy stored on each server 102a, 102b ....102n.

Client computers 118 (hereafter "CC's") associated with members 106 who desire to share or search for shared genetic sequence information are also connected to the Internet 120. The Internet connection allows the client computers to communicate with the GSSS 102. The client computers 118 may be associated with an individual researcher, or with a research establishment 124. In the presently preferred embodiment, each client computer 118 includes client side software 122 (hereafter "CSS"). Client computers 118 used for sharing genetic sequences also may include memory 124 in which genetic sequence information to be shared is stored. A library 126 of such material may be created in the memory 124 to provide a boundary between shared and non-shared files on a member's computer. Such a boundary may be created by using a sub-directory or folder, or may involve the creation of a separate logical drive dedicated to shared information.

In the illustrated embodiment, the CSS 122 allows each CC 118 to either search for or share genetic information, however different versions of the CSS 122 may be made to be limited to searching only, sharing only, or searching and sharing capable. Each

version of CSS 122 which is capable of sharing genetic information includes a formatting tool for converting genetic information on a member's computer into a standardized format, ensuring that other members will be able to access the data.

#### **DataBase and Library Structures**

- 5           As shown in Figure 2, the server side database 110 includes catalog entries 202 for searching for shared biological information. Each unit of biological information shared by a member may be represented by a catalog entry 202a, 202b, ... 204n in the database on the GSSS. The catalog records may also include a flag indicating whether or not a sharing member is presently available on the network. The entries in the database are associated
- 10   with member information records 204a, 204b, ... 204n.

- The biological information 206 stored in a member's library may include a tag name 208 for the information, the biological information itself 210, a chromatograph file 212, research notes 214, journal articles 216 associated with the biological information, and other associated information 218 associated with the shared biological information.
- 15   The information may include multiple files associated with the information, or may utilize a information file which includes one or more of the above associated material.

- The catalog entries 202a, 202b, ... 204n stored in the GSSS database may include the tag name 206 selected by the sharing member to describe the biological information, a reference ID number 220, an ID field for identifying the sharing member 222, an
- 20   accession number 224, the tissue type 226 from which the information was generated, the organism type 228 from which the biological information was derived, and a sequence 230 associated with the biological information.

In addition to the catalog entries, the GSSS database 110 may also contain information associated with the sharing member. This information may include the member ID 222, a reference ID for the member 232, information concerning a member's affiliation 234 with a research organization, reputational information 236 associated with the sharing member, the member's transfer record 238, subscription information 240 associated with the member, and the member's Internet address 242 when the Internet is used as the network communicably connecting the member with the GSSS.

Member records may also be established for member's who do not share any information. The member records are preferably separate records which are linked to catalog entries, allowing a single member record to be associated with multiple catalog entries, reducing the amount of memory required to store the member's records. Alternately, elements of the catalog information may be stored in individual member's libraries where associated shared biological information is stored, further reducing the amount of memory required for the GSSS.

The database 110 of the illustrated embodiment stores cataloging information of genetic sequences as the genetic sequences are shared. The database 110 also maintains a flag 242 associated with shared genetic sequences which indicates whether the shared sequence is presently available for file transfer. The determining factor as to whether a sequence is presently available for file transfer is the existence of an open connection between the GSSS 102 and the CC 118 on which the sequence is stored. When a member logs on to the GSSS 102, the GSSS 102 server updates the database 110 to show that genetic sequences shared by that member 106 are available for transfer. When the member 106 logs off, the GSSS 102 updates the database 110 to show that the genetic

sequences shared by that member 106 are not presently available. This method of segregating shared genetic sequences supports the use of the system as a CC-to-CC transfer system only, since only files that are available to be transferred CC-to-CC get reported to a searching member in a search report.

## 5 Server Side Software

The SSS 108 has several functions. As shown in Figure 1, to accomplish these functions, the software includes several modules, with each module responsible for a given function. A sharing coordinator module 128 interacts with the individual modules to direct the individual modules to accomplish a required task. The individual modules  
10 include a search module 130, a module for receiving new catalog information 132 on newly shared files, indexing 134 the database to reflect file availability, subscribing new members 136, distributing copies of CSS 138, and member authentication modules 140. Other modules, as described below, may be implemented as required.

A principal function of the SSS 112 is to control a repository for references to  
15 shared genetic information stored on individual CC's. As a member 106 places genetic information into a format for sharing, the GSSS 102 is notified by a transmission from the CC 118 to the GSSS 102 of the catalog information for the newly shared files.

The GSSS 102 also maintains a module for administering member subscriptions 136, such as described below. The administration of subscriptions allows the GSSS 102  
20 to maintain access control over the searching capabilities, and over requested transfers of genetic sequence information.

The GSSS 102 may also include an authentication module 140, which allows the GSSS 102 to confirm the validity of a requested CC-to-CC transfer of genetic sequence information. The implementation of an authentication cycle also increases security regarding the accessing of genetic sequence data.

## 5 Client Side Software

As noted above, the client side software (CSS) 122 is responsible for allowing a member to interface with the GSSS 102 through a CC 118, and to facilitate the transfer of information to and from the member's CC 118. To accomplish this, the CSS 122 also includes an interface generating module 146.

10 Figure 3 shows a member graphical user interface display 300 for illustrative purposes. The interface has four primary function groups which a member may access through the illustrated interface. These are a chat 302 function, a set of library functions 304, a search function 306, and a data transfer function 308. Each function may be invoked by clicking on the tab representing each function.

15 The chat function 302 allows a member to enter into a conversation with another member over the Internet. The presently illustrated embodiment uses a typed chat function, wherein the members are able to enter messages using the keyboard of their associated client computer 118. The chat function may also be implemented by Internet telephony, or Internet videoconferencing.

20 The chat function allows researchers to correspond regarding shared genetic sequence information. For example, a researcher who has identified what he or she believes to be a relevant genetic sequence may have questions regarding how the sequence

was derived. The researcher may contact the member who has shared the data to inquire how the information was derived. Since the database 110 already has implemented a signal regarding the availability of the sharing member, the availability flag may also be used to only present a sharing member's contact address when the sharing member is  
5 connected to the GSSS 102.

The library function 304 allows a member to set up a library 126 for storing genetic sequences which the member is willing to share, assists the member in creating formatted copies of the genetic sequences which are to be shared, assists the member in providing cataloging information regarding genetic sequences to be shared, and allows the  
10 member to remove a shared genetic sequence from the library, thus removing its availability for sharing.

The searching function 306 allows a member to define parameters for a genetic sequence which the member is searching for. The search may be for a specific sequence string, and may also may be further refined based on other search criteria. Search criteria  
15 illustrated in Figure 2 include gene name 312, member code 314, accession number, organism, and tissue type. The member code may be used to search for shared genetic sequences shared by a specific researcher. Other search parameters may include, but are not limited to, dates when a genetic sequence was initially shared and researcher type information, such as whether the source of the shared file is a private or public  
20 organization, and whether the source of the genetic sequence places additional conditions on the use of the genetic sequence.

The presently preferred search method compares a sought after sequence to

sequences associated with shared biological information. A searching member may specify a tolerance for the search, where the tolerance is based on the number of positions in a sequence which differ from the sought after sequence. The tolerance may be described in a percentage of positions which match the search criteria.

- 5 Searches may be performed using the GSSS database only, the GSSS database in conjunction with direct review of shared material, or direct review of shared material only. In the case of database only searches, the GSSS reviews a member's search parameters against the catalog information stored in the database, and returns hits to the member based on the limited information stored in the database. This may result in less than  
10 optimal search results, as additional characteristics of the shared information may not be apparent from the catalog information.

- Direct review of shared information may utilize the transmission of the search parameters to sharing member's computers, such that the CSS installed on the machine may perform a comparison of the parameters to the material shared in the member's  
15 library. As shown in Figure 2A, the sequence and accession number may be stored in a member's library, reducing the amount of information which needs to be stored in the database 110. This method allows for richer searching, as more information than can be efficiently stored on the GSSS or database can be examined, using the distributed processing power of each sharing member's computer. Once a CC has compared shared  
20 biological information to the search parameters, the CC reports to the GSSS whether any material meeting the search parameters is stored on the computer, and if so, the identity of the shared information which meets the parameters. Alternately, the CC may report only positive results to reduce the amount of information transmitted to the GSSS.



The search method may also be a combination of database and direct searching, utilizing the database to narrow the searched files or libraries, and thus compromising on the required data transmissions while still allowing the richer direct comparison results.

The search parameters for genetic sequences may include a known string of values, wherein a member seeks genetic sequences containing the string. In this case, a searching member may also tolerate a threshold level of matching, such as, for example, that 70% or 80% of the values in the string are matched in a shared file. This threshold value may be specified by a searching member, and used in conjunction with other search parameters to narrow or broaden a search.

The transfer function 308 is shown in Figure 4 as embodied in a graphical user interface between a member and the member's associated client computer. The transfer interface 400 may include, but is not limited to, a display of the catalog information 402 associated with the source of a shared file being transferred, a summary value for reputational information 404, and statistical information 406 regarding the searching member's activities.

A genetic sequence to be downloaded may appear in a Sequence Alignment Window 408 along with any overlapping sequences that are present in the searching member's library 410, 412 [Sequence 1 and Sequence 2 in figure]. These genetic sequences appear in a window along with a copy of the search string 414. The search string 416 [represented by the black bar in the diagram] is displayed within the context of the entire sequence to show the member the direct area of overlap. The similarity between the search string and the resultant sequence may be expressed with a separate value 418.

Sequence 2 (element number 412) represents an additional sequence in the member's library which overlaps with the downloaded sequence but not with any other sequence in the library. In this way, the member is able to build a larger contiguous genetic sequence (e.g., the member has sequence information spanning base pairs 200-1500, instead of just  
5 200-1322 if Sequence 2 is absent).

The sequence overlap presentation may also be dependant on a tolerance defined by the searching member, such that individual files stored in a member's library may be compared to the biological information being considered for transfer, with each file defining biological information in the sharing member's library being compared to the  
10 information being considered, and an overlap value determined for each comparison between a file being considered for transfer and an individual file in the searching member's library.

After reviewing the sequence specifications, member specifications and viewing the alignment of the sequence as it compares to the sequences in the member's library, the  
15 searching member will have the option of transferring a copy of the sequence information from the sharing member's CC to their own library by selecting a complete transfer button 420. If the searching member is not satisfied with the sequence information, the searching member may decide to cancel the transfer by selection cancel transfer button 422.

A library maintenance function associated with client side software may be  
20 implemented to allow a member to add biological information files to the member's library. This addition function can designate information as either shared or closed. Files that a member does not desire to share may be designated as closed to allow the member

to designate files for comparison purposes during search and transfer functions.

In addition to the express functions provided on the interface, the CSS 122 may also act as a firewall between members and the GSSS 102 on the Internet 104. Such a firewall is useful to prevent members from providing information for sharing which may be harmful to other members, or to the GSSS 102. The illustrated embodiment includes a computer-virus scanning routing 148 (Shown in Figure 1) which is executed when genetic sequence files are formatted for sharing. The scanning routine works simultaneously with the formatting routine, such that any genetic sequence that is reformatted to be shared is also virus-scanned. If the scanner 148 detects the potential presence of a computer virus in the file, the scanner 148 may mark the file as potentially infected, and report the condition to the GSSS 102. References to a shared file which is marked as potentially infected will not be added to the database, thus preventing the possibly infected file from being unwisely transferred. Additional actions may also be initiated upon discovery of a potentially infected file, such as quarantining the member who attempted to share the file, and/or requiring all shared files in the member's library to be rescanned.

#### **Member Authentication**

A member authentication routine 150 may also be employed to ensure that copy transfer requests, when the requests are issued client-to-client, are valid requests from members of the network of members. The member authentication check 150 functions by verifying that the received transfer request is from a member who received the address by conducting a valid search utilizing the GSSS 102. Each time a member runs a search, the GSSS 102 generates the list of shared genetic sequences that meet the searcher's criteria.

The member selects a link to a sharing member's client computer that houses the shared genetic sequence. By logging each link that the searching member actuates, the OSSS 102 creates a log of valid transfers created by the GSSS 102.

5 A transfer request made to a sharing member's could be made without the use of the GSSS 102, and thus be an invalid transfer request. In order to limit the potential for such invalid transfer requests, the sharing member's CC 118 may query the GSSS 102 as to whether the received transfer order was generated through the GSSS 102. If the request was not made through the GSSS 102, the sharing member's CC 118 may be so notified, and instructed to refuse the transfer request.

10 An alternate embodiment of the authentication routine may rely on the use of certificates as a means of verifying that the searching member is a member of the genetic sequence sharing network. Certificates are identification data blocks issued by certificate authorities, where the data blocks may be encrypted using asymmetrical encryption based on private-key, public key methods. The client side software may either decrypt the  
15 certificate itself, or reflect the certificate to the GSSS 102 for the GSSS 102 to decrypt. If the certificate is not valid, the transfer may be refused.

#### **Peer Review and Chromatogram Options**

The present system allows genetic information to be shared without requiring the rigor of scrutiny before the data is shared. Present methods of publicly publishing genetic  
20 information require the data to be scrutinized before it is published, for academic and liability reasons. As such, a researcher must first format the data in a form acceptable to a reviewing body, answer questions regarding the derivation of the data, and then await a

publication decision from a publishing board. This distinction allows genetic information to be published without requiring a researcher who derived the data to expend resources to publish the information, thus allowing more genetic information to be made available for the same level of effort. However, the lack of scrutiny of the genetic sequence being

5 shared results in some limitations as to the value of the genetic sequence.

In order to provide an indicator of the accuracy of the genetic sequence, two options may be integrated with the genetic sequence transferring system. The first is a peer review reference, wherein comments regarding the accuracy or inaccuracy of a specific member's shared genetic sequences may be received, stored, and then indexed for

10 review in association with all of the shared genetic sequences of that specific member. The second method may be to provide a chromatograph image file from which a genetic sequence is derived along with the shared genetic sequence file itself during transfer. The second method may be expanded to include additional reference files with the shared information, such as articles written regarding the biological information, chromatograph or

15 other image files, files containing researcher notes regarding the shared information, or other files which may be of use to a searching member. For ease of description, shared information and/or a shared file may hereinafter refer to either the biological information itself or the joined biological information and associated files.

A chromatograph is a graphical representation of raw sequencer results. The

20 chromatograph allows a researcher to compare a given genetic sequence against the sequence's underlying chromatograph to review the resultant sequence data, in effect is double-checking the derived genetic sequence.

As a second means for providing an indication of the reliability of a shared genetic sequence, a copy of the chromatograph from which the sequence was derived may be attached to a transferable or transferred genetic sequence. Since it is difficult to falsify chromatographs, the attachment of the chromatograph to a sequence file provides a measure of authentication of the validity and accuracy of the genetic sequence.

#### **Data Transmission**

The transmission of genetic information files from one member's computer to another member's computer preferably takes place between the two computers as a means of reducing traffic through the GSSS 102. By utilizing direct transfers between client computers, only search criteria, search results, and authentication information must be transmitted by, to, or through the GSSS 102, thus reducing the bandwidth required to support the network of members.

Since some degree of confidentiality or proprietary control may need to be exercised over shared genetic sequences, such as in the case where the sequence is owned by a private company, the client side software may include encryption and digital rights management technologies.

A variety of encryption technologies may be implemented, however the system of the present illustration is based on a limited public-key encryption system such that the client side software may be distributed to researchers in foreign countries without breaching encryption export limitations. If the network of researchers is intended to be entirely within the United States, higher levels of encryption may be utilized.

The encryption algorithm encrypts data being transferred from a sharing member's

computer, and another member's client side software decrypts the transferred data when received. In order to accomplish this, the client soft software modules may use secure socket layer protocols, transport layer security protocols, and/or other forms of public-key encryption.

- 5           Also, digital rights management techniques may be incorporated into the genetic sequence files, such that access to the genetic sequence, and/or the ability of a member to copy or further share a genetic sequence, is controlled by a member who shares the original genetic sequence. The implementation of such technology allows a member to control access to a shared genetic code, even after the code has been transferred to a
- 10   searching member. This allows private owners of genetic sequences, who desire to charge for access to the sequence, to limit a transferal of a genetic sequence to only the recipient of the sequence. Also, embedded icons can be inserted into shared biological information to determine whether a file has been modified since it was originally transferred, and also who the original source of the file was. Since either the GSSS 102, or the sharing
- 15   member's computer, may track the transferal of genetic sequence files, the transfers may be also tracked for billing purposes. Transfer records documenting individual file transfers, may also be associated with individual members, tracking both transfers to and transfers from, to allow further historical analysis of shared biological information.

- Implementing the tracking service into the GSSS 102 allows small researchers
- 20   who develop genetic sequence information to post such information without first having to generate tracking and billing infrastructure to support a pay-per-share posting of the genetic sequence.

**Delayed Transmission**

In an alternate embodiment of the present invention, the GSSS 102 may provide support for the transfer of shared genetic sequences from sharing member's CC's 118 when the CC's 118 are not connected to the Internet. Such a delayed transfer may be accomplished by the GSSS 102 presenting all files that meet a searching member's criteria, not just those belonging to a presently connected member. When a searching member requests transfer of a shared genetic sequence from a sharing member that is not presently connected, the GSSS 102 may generate a file transfer request to the sharing member, with the GSSS 102 as the recipient of the file transfer, followed by transfer of the file from the GSSS 102 to the searching member when the searching member is connected to the GSSS 102. Such an implementation requires the GSSS 102 to maintain available memory for the temporary storage of delayed transfer files. In the described embodiment, transfer requests from sharing members that are connected to the GSSS 102 would occur CC-to-CC, as utilized in the embodiment described above. A searching member could apply sequence comparison tools as described above, allowing a right of rejection of a file that differs from what the searching member believed himself to be requesting.

**Method for Providing a Genetic Information Sharing Service**

The above system may be implemented in a business activity which facilitates the sharing of genetic information. Such an activity is premised on the providing 502a GSSS 102 as described above. With the server in operation, the business activity may subscribe users 504 to the sharing service. The subscriptions may be fee based or free.



Fee based subscriptions may also accommodate varying rates, with discounted rates provided for members who contribute shared genetic information, while members who only search for genetic information are required to pay a full rate. In the case of a fee-based system, the subscription may be a flat-rate subscription, or may implement  
5 different rates based on usage.

Usage-based subscription rates may be based on a fixed amount per search request. Additionally, free searches could be provided to a member in exchange for the member making a unit of genetic information available for sharing. The usage-based subscription rate may also be combined with a fixed fee, such that upon payment of the fixed fee, a set  
10 number of searches are free to the member. Such a hybrid system may also be implemented by varying the fixed portion of a hybrid fixed/usage fee inversely with a per-search rate. As such, a member who desires to make frequent searches may pay a high fixed fee, but a low per-search fee, resulting in lower overall costs.

Once members have been subscribed to the system, client side software is  
15 distributed 506. The client side software provides a member with the ability to format genetic information for sharing, and to provide information related to cataloging the shared genetic information. The software distribution may either be by physical delivery of a storage medium, or by providing a download site for a member to download the software from. Once the member has the copy of the client side software, the member  
20 may be instructed to install 508 the software on the client computer associated with the member.

The software distribution may also involve the distribution of a security such as a password, certificate, or encryption key. The security code may also enable the client side software. Such a security code may also be distributed from the client side software.

Once a member has installed the software on the client computer associated with  
5 the member, the member may be prompted to format 512 and provide cataloging data for information that the member has a willingness to share. Once the data is formatted and the cataloging information has been provided, the client computer, under the direction of the client side software, reports 514 the cataloging information and identity of the member to the GSSS 102.

10 The information sharing receiver receives 516 the information, and stores 518 the information in a database. The GSSS 102 also may associate a link 520 to the member's CC 118 with the genetic information, such that later retrieval of the cataloged information by a searcher results in the GSSS 102 being able to identify the path to the shared data. Also, the GSSS 102 may associate a reference 522 to reputational information associated  
15 with the member to the cataloged information, such that members may consider the reputational information of the source of genetic information identified on a search before requesting a copy of the genetic information.

In the presently described embodiment, files may only be shared while a sharing member is connected to the GSSS 102. When the GSSS 102 receives cataloging  
20 information from a sharing member, the GSSS 102 indexes 524 the database to reflect the availability of the information for sharing. The availability of information for sharing may be indicated by the setting of a flag associated with a member or a file. When a

member logs off, the server is notified of the unavailability of the information by the log off procedure, and updates the database to remove files shared by the member who logged off.

Also in the presently described embodiment, it is envisioned that a plurality of  
5 servers may be employed to meet member demand. Each server in such a configuration contains a copy of the database for searching. In this configuration, one server is designated as the master server, and is responsible for re-indexing the data-bases maintained by each server, and publishing 526 the re-indexed information to the other servers.

10 Also, the system may maintain the catalog information in memory when a member logs off and only remove a reference signifying the availability of the shared files, such that a later log-on only requires entry of new shared files, and the re-setting of a flag indicating that a member is connected to the network. The system may also send test communications to members to determine if they are still connected to the network in  
15 advance of file transfer requests, or may merely monitor transfer requests for failed transfer requests, and update the availability information in the database based on a communications failure.

An alternate embodiment of the system may rely on the use of standard Internet browsers for viewing generated display pages on the GSSS 102, obviating the need to  
20 distribute client side software to members who only intend to search. Once a member has conducted a search, and identified desired genetic information, the information sharing software may issue an electronic request to the client computer on which the information

is shared for the client computer to transmit a copy of the genetic information directly to the requesting member, for example using FTP (file transfer protocol) or HTTP (hypertext transfer protocol).

Once the system has received 528 a search request from a member, the GSSS 102  
5 acquires parameters 530 describing the genetic information sought by a member. This information is compared 532 to the catalog information in the database. A list of shared information which meets the search criteria is generated 534 for the searching member, and transmitted 536 to the searching member's computer.

If it is determined 538 that the searching member desires to further examine shared  
10 information listed in the search results, as shown in Figure 3, the user may identify information for further examination by selecting the information on the search list. Selecting the information on the search list transfers 540 the searching member to the transfer graphical user interface, shown in Figure 3. The GSSS 102 in conjunction with the client side software may also generate 542 a sequence comparison display as shown in  
15 Figure 4. The searching member may then decide to transfer the file, to examine a different file, or to conduct a new search. If it is determined 546 that the searching member desires to begin a new search, the process reverts to step 528. If it is determined that the member desires to view further information on different shared information, the process reverts to step 536.

20 If it is determined 544 that a searching member desires to obtain a copy of shared information identified as meeting that member's search criteria, the searching member merely selects the address link for the desired shared information from the list of shared

information meeting the member's search criteria. Once it has been determined 540 that a searching member desires reputational information, a list of reputational information associated with a source of shared information is generated 542 and transmitted to the searching member.

- 5           In a usage based system, the location of the shared information may be implemented as a blind link to the location, such that a member must use the link to identify the destination. Such a system would allow a GSSS 102 to base usage fees on a per transfer basis, rather than on a per search or fiat fee basis. The use of blind links would also allow copy fees for a sharing member to be collected if required by a sharing
- 10   member.

- The actual transfer 546 of the file may be implemented by peer-to-peer sharing, wherein the file is transferred from the sharing member's computer to the searching member's computer directly, or by transferring the file to the GSSS 102, and from there to the searching member's computer. Alternately, a hybrid system may also be
- 15   implemented, where shared information from a member who is presently connected may be transferred peer-to-peer, while delayed file requests (because the sharing member is not presently connected to the net) may be handled by uploading the file to the GSSS 102 when the sharing client is available, then transferred to the searching member's computer when the searching member is available.

- 20           When peer-to-peer transfers are requested, the system may incorporate an member authentication procedure to confirm for a sharing member that the persons requesting the genetic information is authorized to receive the material. The sharing member's computer

receives the request, and identifies the requestor. The sharing member's computer then transmits an authorization or member validation query, including the identity of the requestor, to the GSSS 102. If the GSSS 102 confirms that the requestor is a valid member, then the GSSS 102 transmits an approval or authorization to the sharing  
5 member's computer. The inclusion of the authorization check provides a redundant means for the GSSS 102 to track the genetic information file transfer requests made by a searching member for fee calculation purposes.

### **Server-less Genetic Information Sharing**

In an alternate embodiment, the genetic information sharing system may be  
10 propagated between peer computers, wherein search routines are based on echoed queries from member to member of a pool of member's who desire to share genetic information. As opposed to the client server implementation described above, the peer-to-peer implementation relies only on software agents installed on member's computers.

In addition to the above-described function of formatting genetic information to be  
15 shared; the software agent is responsible for cataloging the information available to be shared on the member computer, and searching the cataloged information in accordance with a search request. The software agent may store cataloged information prior to receiving a search request, or may store the catalog information in individual shared information files, and search the files upon a search request. Additionally, the software  
20 agent is capable of forwarding the search request to other computers known to be part of the network, and to have the software installed thereon.

When a member wants to search for genetic information, he generates a search

request on his or her computer. Next, he or she determines the address of at least one other computer that is part of a network of computers associated with members who are involved in the sharing of genetic information. This determination may be accomplished by the member's computer connecting to a known Internet location to obtain the address of an associated computer that is presently connected to the Internet, or by reference to a previously known address or group of addresses.

The search request is then transmitted via the Internet to the at least one other computer address. Since the computer located at the address to which the transmission was sent is responsible for searching itself for shared information meeting the search criteria and forwarding the request to other computers associated with the sharing group, the initial transmission must be sent to a computer that is turned on and connected to the Internet. The initial transmission of the search request to multiple addresses will increase the likelihood of the message being transmitted to an active computer.

Each computer receiving the message searches itself for shared information meeting the search criteria. If the receiving computer determines that it has a file or files containing information which meets the search criteria, the receiving computer may generate a message to the search computer that the receiving computer has files meeting the search criteria, or alternately send the files meeting the search criteria.

If the search computer does not have shared information that meets the search criteria, the receiving computer forwards the search request to at least one other computer associated with the genetic information sharing association.

Search requests may be controlled in several ways. A simple time-out counter

may be included in the original search request. Each time a computer searches for shared information, the counter is incremented. Once a search request has been forwarded a predetermined number of times, the receiving computer will search its own files for shared genetic information that meets the search criteria, and then dispose of the message  
5 without forwarding it.

Alternately, the search request may be forwarded from a receiving computer to only a single forwarding address, wherein a receiving computer that has shared genetic information which meets the search criteria transmits the shared material to the searching member's computer, then discards the search message.

10 A further embodiment of the system for sharing biological information may be implemented using a hybrid system that uses a server for storing addresses for member's who have biological information available for sharing. The hybrid system thus only stores member information in the server side database.

Search routines using a hybrid system can either rely on client side software to  
15 perform searches on shared information, or may use server side software to search libraries on member's computers to which the GSSS has been granted access authority. Client side software for this implementation receives a search request from the GSSS, and then searches libraries contained on the member computer for information which meets the search criteria. If shared information meeting the criteria is found, the member  
20 computer reports the presence of the information to the GSSS, which then compiles a list of information meeting the criteria for the searching member.

The hybrid system can also use the GSSS to perform searches on information



stored on individual member's computers. This system relies on access to the shared information on member's computers. The GSSS uses this authority to access the shared information, and search the information for information meeting a set of search criteria. Such a search may be accomplished by transferring from the GSSS to the member  
5 computers a self-executing program, such as a Java applet or web spider, to shared libraries. The executable program searches information stored in the library for information meeting search criteria, then reports to the GSSS when information meeting the search criteria is found. Once the self-executing program reports back to the GSSS regarding discovered information, or if no information meeting the criteria identified, the  
10 self-executing program may delete itself.

From the above, it is evident that the present invention may be embodied in other specific forms without departing from the spirit or essential attributes of the invention. Accordingly, reference should be made to the appended claims, rather than the foregoing specification, as indicating the scope of the invention.

### Claims

What is claimed is:

1. A system for sharing biological information utilizing a network, comprising:

a plurality of subscribing members, said subscribing members each having a network device for accessing a network, wherein at least one subscribing member has biological information for sharing, said biological information stored in memory on the subscribing member's network device;

a network host, said network host having a database,

said database containing records identifying biological information stored on member's computers, each of said records containing at least one descriptive element selected from a group of descriptive elements consisting of sequence data, a tissue source descriptor, or an organism type descriptor, and information describing an identity of a network device of a subscribing member wherein biological information associated with said record is stored;

wherein the plurality of subscribing members and the network host are communicably interconnected by a network.

2. A system for sharing biological information according to claim 1, wherein the subscribing member network devices are computers, and further include client side software, said client side software for assisting a subscribing member in formatting biological information to be shared into an acceptable format for sharing.

3. A system for sharing biological information according to claim 1, wherein the subscribing member devices are computers, and further include client side software, said client side software system for sharing biological information for computer virus infection.

4. A system for sharing biological information according to claim 1, wherein network devices are computers, and further include client side software, for encrypting biological information for sharing.

5. A system for sharing biological information according to claim 1, wherein the subscribing member network devices are computers, and further include client side software, said client side software for employing digital rights management technology to limit re-sharing of a shared biological information file.

6. A system for sharing biological information according to claim 1, wherein the biological information stored on a subscribing member's network device is a file related to a genetic sequence, said sequence comprising a plurality of position values associated with sequence values.

7. A system for sharing biological information according to claim 6, wherein the client side software further is capable of associating an image file with shared biological information, and sharing the associated image file with the biological information.

8. A system according to claim 1, wherein the network host further includes biological information sharing software, said software for receiving information from a subscribing member identifying shared biological information on said subscribing member's network device, said software further allowing information stored in said database to be searched relative to criteria provided by a subscribing member.

9. A system for sharing biological information according to claim 8, wherein the biological information sharing software further includes functionality for receiving reputational information associated with a member who is sharing information and associating the received reputational information with biological information shared by that member.

10. A system for sharing biological information according to claim 8, wherein the biological information sharing software further includes functionality for receiving a request from a sharing member to authenticate a request to share biological information from another member of the network.

11. A system for sharing biological information according to claim 8, wherein the biological information sharing software further is capable of recording transfers requested by individual subscribing members.

12. A system for sharing biological information according to claim 1, wherein the network that communicably interconnects the subscribing members and the network host is the Internet.

13. A system for sharing biological information according to claim 12 wherein the network host includes a plurality of server computers, each server computer containing biological information sharing software.

14. A system for sharing biological information, the system comprising:  
a plurality of subscribing members;  
a means for disseminating information regarding the location of stored biological information, said means for disseminating information capable of receiving

information from the plurality of subscribers identifying where biological information is stored, providing a database, said database including records describing stored biological information, said descriptions containing at least one descriptive element selected from a group of descriptive elements consisting of sequence data, a tissue source descriptor, or an organism type descriptor, said records further referring subscribing members to the storage location;

at least one means for said plurality of subscribers to communicably connect to the means for disseminating information, wherein said at least one means for subscribers to communicably connect to the means for disseminating information includes a memory for storing biological information.

15. A system according to claim 14, wherein the means for disseminating information further is capable of receiving search criteria from a member, and comparing said search criteria to the information received from the plurality of members to determine if biological information stored on a member's network device meets the search criteria.

16. A system according to claim 15, wherein the means for disseminating information further is capable of generating a list of biological information stored on network devices, said list including only biological information that meets search criteria.

17. A system according to claim 14, wherein the at least one means for said plurality of subscribers to communicably connect to the means for disseminating information further is capable of receiving from a subscriber a set of criteria, said criteria defining biological information sought by the member.

18. A system according to claim 17, wherein the criteria includes at least one sequence value.

19. A system according to claim 17, wherein the parameters include an accuracy threshold, said accuracy threshold defining a number of sequence values which can differ from a specified set of sequence values used as a criteria.

20. A method for exchanging biological information, comprising the steps of:  
providing a host computer, said host computer being connected to a network;

subscribing a plurality of members, said members each having a network device capable of accessing the host computer via a network;

providing a database on said host computer, said database for storing descriptive elements describing biological information stored on member's network devices and for storing locations associated with the biological information stored on member's network devices, said descriptive elements containing at least one descriptive element selected from a group of descriptive elements consisting of sequence data, a tissue source descriptor, or an organism type descriptor;

receiving via the network descriptive elements from members describing biological information stored on a member's network device;

storing the received descriptive elements in the database;

receiving from a member a request for biological information, said request including criteria describing biological information sought by the requesting member, said criteria including at least one sequence value;

comparing the request criteria with descriptive elements stored in the database;

informing the requesting member of the stored location of biological information when criteria stored in the database meets the request criteria.

21. A method for exchanging biological information according to claim 20, further comprising the step of providing the plurality of members with a computer program, said computer program for installation on the member's network device, said computer program for assisting the member in sharing biological information.

22. A method for exchanging biological information according to claim 20, further comprising the step of indexing the database to indicate when individual members are connected to the network such that biological information may be transferred from one member's network device to another's.

23. A method for exchanging biological information according to claim 22, further comprising the steps of receiving from a searching member a request for the host computer to determine whether biological information described in the database meets search criteria provided by the searching member; determining whether biological information described in the database meets the search criteria provided by the searching member; and informing the searching member of the stored descriptive elements describing biological information which meets the search criteria.

24. A method for exchanging biological information according to claim 23, wherein the step of informing the searching member of the presence of the biological information which meets the search criteria includes generating a list of individual

biological information references, wherein each individual biological information reference describes a single unit of biological information stored on a member's network device.

25. A method for exchanging biological information according to claim 24, further comprising the steps of:

receiving from a searching member an indication that the searching member would like to receive a copy of a unit of biological information identified in the list;

determining whether the network device on which the desired biological information is stored is communicably connected to the network host;

when the network device on which the desired biological information is stored is communicably connected to the network host, transmitting to the network device on which the desired biological information is stored a request for a copy of the indicated unit of biological information to be transferred to the searching member.

26. A method for sharing biological information according to claim 25, further comprising the steps of:

when the network device on which the desired biological information is stored is not communicably connected to the network host, storing a request to transfer a copy of the desired biological information;

determining when the network device on which the desired biological information is stored becomes communicably connected to the network host;



when it is determined that the network device on which the desired biological information is stored is communicably connected to the network host, determining whether the searching member is communicably connected to the network host;

when it is determined that both the network device on which the desired biological information is stored and the searching member are communicably connected to the network host, transmitting to the network device on which the desired biological information is stored the stored request to transfer a copy of the desired biological information to the searching member.

27. A method for sharing biological information according to claim 26, further comprising the steps of:

when it is determined that the network device on which the desired biological information is stored is communicably connected to the network host, determining whether the searching member is communicably connected to the network host;

when it is determined that the network device on which the desired biological information is stored is communicably connected and that the searching member is not communicably connected to the network host, transmitting to the network device on which the desired biological information is stored the stored request to transfer a copy of the desired biological information to the network host;

storing the desired biological information on the network host; and

transmitting the desired biological information from the network host to the searching member when the searching member is next communicably connected to the network host.

28. A method for exchanging biological information according to claim 22, further comprising the steps of receiving from a searching member a request for the host computer to determine whether biological information described in the database meets search criteria defined by the searching member; transmitting the search criteria to member's computers on which biological information is stored, receiving comparison results from member's computer's identifying shared biological information which meets the search criteria provided by the searching member; and informing the searching member of biological information having descriptive elements which meet the search criteria.

29. A computer-readable medium tangibly embodying instructions which, when executed by a computer, implement a process comprising the steps of:

prompting a member of a computer executing the computer-readable medium to identify biological information that the member is willing to share;

prompting a member to define criteria describing biological information that the member is willing to share;

formatting biological information that the member is willing to share into a standardized format;

storing the formatted biological information in a known location on the computer executing the computer-readable medium;

transmitting to another computer the criteria describing the shared biological information and the location on the computer where the formatted biological information is stored.

30. A computer-readable medium according to claim 29, wherein said instructions when executed by a computer implement a process further comprising the step of scanning biological information to be formatted to detect if the biological information may be infected with a computer virus.

31. A computer-readable medium according to claim 30, wherein said instructions when executed by a computer implement a process further comprising the step of quarantining biological information to be shared when a computer virus is detected.

32. A computer-readable medium according to claim 29, wherein the another computer is a network host, and wherein the instructions implement a process further comprising the step of requesting an authentication from the network host when a request to share biological information is received from another computer.

33. A computer readable medium according to claim 29, wherein said instructions when executed by a computer implement a process further comprising the step of transmitting to another member a copy of shared biological information when the another member requests a copy of the shared biological information.

34. A computer-readable medium according to claim 33, wherein said instructions when executed by a computer implement a process further comprising the step of transmitting a chromatograph associated with shared biological information when said shared biological information is transmitted to the another member.

35. A computer-readable medium according to claim 33, wherein said instructions when executed by a computer implement a process further comprising the step of encrypting the copy of shared biological information before transmitting the copy to the another member.

36. A computer-readable medium according to claim 29, wherein the another computer is a network host, and wherein said instructions when executed by a computer implement a process further comprising the steps of:

prompting a member to provide search criteria when the member desires to search for shared biological information;

transmitting the search criteria to the network host;

receiving from the network computer references to locations where shared biological information meeting the provided search criteria is stored.

37. A computer-readable medium according to claim 36, wherein said instructions implement a process further comprising the step of transmitting a request to a location where biological information meeting the provided search criteria is stored, said request requesting that the recipient of the request transmit to the sender of the request a copy of the shared biological information which meets search parameters.

38. A computer-readable medium according to claim 37, wherein said instructions implement a process further comprising the step of receiving biological information which meets search criteria.

39. A computer-readable medium according to claim 38, wherein said instructions implement a process further comprising the step of decrypting received biological information when the received biological information has been encrypted by a sender.

40. A computer-readable medium according to claim 38, wherein said search criteria comprise a genetic sequence, wherein said instructions implement a process further comprising the steps of:

determining whether biological information stored on the searching member's computer has a genetic sequence matching the genetic sequence searched for;

when biological information on the searching member's computer has a genetic sequence matching the genetic sequence searched for, displaying the received biological information in a format allowing a member to compare the position of the genetic sequence of the biological information stored on the computer to the biological information received; and

when the member of the computer determines, after comparing the positioning of the genetic sequence in the received biological information to the position of the genetic sequence of biological information stored on the searching member's computer, that the member does not desire to retain the received biological information, discarding the biological information and informing the network host that the biological information has not been retained.

41. A business method for providing a biological information exchange, comprising the steps of:

providing at least one host computer, said at least one host computer being connected to a network;

subscribing a plurality of members, said members each having a computer capable of accessing the host computer via the Internet;

providing a database on said at least one host computer, said database for storing criteria describing biological information stored on member's computers and for storing locations associated with the biological information stored on member's computers, said criteria including at least one descriptive element selected from a group of descriptive elements consisting of sequence data, a tissue source descriptor, an organism type descriptor, or an accession number,

receiving via the Internet descriptive elements information from member's describing biological information stored on a member's network device;

storing the received descriptive elements in the database;

receiving from a member via the Internet a request for biological information, said request including criteria describing biological information sought by the requesting member;

comparing the request criteria with descriptive elements stored in the database;

informing the requesting member of the presence of stored biological information stored on another member's computer when descriptive elements stored in the database meets the request criteria.

42. A business method for providing a biological information exchange according to claim 41, further comprising assessing each member a subscription fee for subscribing to the biological information exchange.

43. A business method for providing a biological information exchange according to claim 41, further comprising the steps of providing the plurality of members with a computer program for a fee, said computer program for installation on the member's network device, and assisting the members in sharing biological information.

44. A business method for providing a biological information exchange according to claim 41, further comprising the steps of providing the plurality of members with a computer program, said computer program for installation on the member's network device, and assisting the members in sharing biological information, and indexing the database to indicate when individual member's are connected to the network such that biological information may be transferred from one member's network device to another's.

45. A process for exchanging biological information according to claim 44, further comprising the steps of receiving from a searching member a request for the host computer to determine whether biological information described in the database meets search criteria defined by the searching member; determining for a fee whether biological information described in the database meets the search criteria defined by the searching member; and informing the searching member of biological information having descriptive elements which meet the search criteria.

46. A process for exchanging biological information according to claim 44, further comprising the steps of receiving from a searching member a request for the host computer to determine whether biological information described in the database meets search criteria defined by the searching member; determining whether biological information described in the database meets the search criteria defined by the searching

member; and informing the searching member of the stored criteria describing biological information which meets the search criteria, wherein the step of informing the searching member of the presence of the biological information which meets the search criteria includes generating a list of individual biological information references, wherein each individual biological information reference describes a single unit of biological information stored on a member's network device.

47. A process for exchanging biological information according to claim 45, further comprising the steps of:

receiving from a searching member an indication that the searching member would like to receive a copy of a unit of biological information identified in the list;

charging the searching member a fee associated with the request to receive a copy of a unit of biological information identified in the list;

determining whether the network device on which the desired biological information is stored is communicably connected to the network host;

when the network device on which the desired biological information is stored is communicably connected to the network host, transmitting to the computer on which the desired biological information is stored a request for a copy of the indicated unit of biological information to be transferred to the searching member.

48. A process for sharing biological information according to claim 47, further comprising the steps of:



when the network device on which the desired biological information is stored is not communicably connected to the network host, storing a request to transfer a copy of the desired biological information;

determining when the network device on which the desired biological information is stored becomes communicably connected to the network host;

when it is determined that the network device on which the desired biological information is stored is communicably connected to the network host, determining whether the searching member is communicably connected to the network host;

when it is determined that both the network device on which the desired biological information is stored and the searching member are communicably connected to the network host, transmitting to the network device on which the desired biological information is stored the stored request to transfer a copy of the desired biological information to the searching member.

49. A process for sharing biological information according to claim 48, further comprising the steps of:

when it is determined that the network device on which the desired biological information is stored is communicably connected to the network host, determining whether the searching member is communicably connected to the network host;

when it is determined that the network device on which the desired biological information is stored is communicably connected and that the searching

member is not communicably connected to the network host, transmitting to the network device on which the desired biological information is stored the stored request to transfer a copy of the desired biological information to the network host;

storing the desired biological information on the network host; and

transmitting the desired biological information from the network host to the searching member when the searching member is next communicably connected to the network host.

50. A process for exchanging biological information according to claim 46, further comprising the steps of:

receiving from a searching member an indication that the searching member would like to receive a copy of a unit of biological information identified in the list;

charging the searching member a fee associated with the request to receive a copy of a unit of biological information identified in the list, wherein the fee is dependent upon the number of requests to receive a copy of a unit of biological information made by the searching member;

determining whether the network device on which the desired biological information is stored is communicably connected to the network host;

when the network device on which the desired biological information is stored is communicably connected to the network host, transmitting to the computer on which the desired biological information is stored a request for a copy of the indicated unit of biological information to be transferred to the searching member.

51. A process for sharing biological information according to claim 50, further comprising the steps of:

when the network device on which the desired biological information is stored is not communicably connected to the network host, storing a request to transfer a copy of the desired biological information;

determining when the network device on which the desired biological information is stored becomes communicably connected to the network host;

when it is determined that the network device on which the desired biological information is stored is communicably connected to the network host, determining whether the searching member is communicably connected to the network host;

when it is determined that both the network device on which the desired biological information is stored and the searching member are communicably connected to the network host, transmitting to the network device on which the desired biological information is stored the stored request to transfer a copy of the desired biological information to the searching member.

52. A process for sharing biological information according to claim 50, further comprising the steps of:

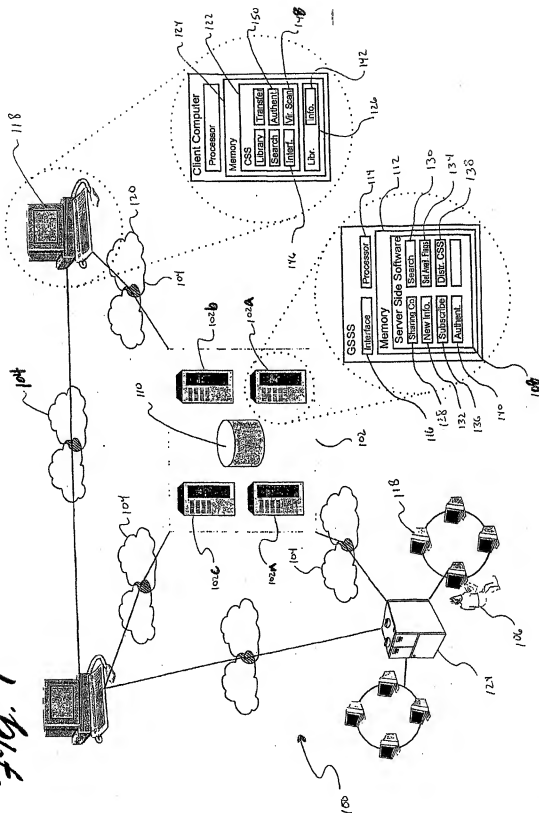
when it is determined that the network device on which the desired biological information is stored is communicably connected to the network host, determining whether the searching member is communicably connected to the network host;

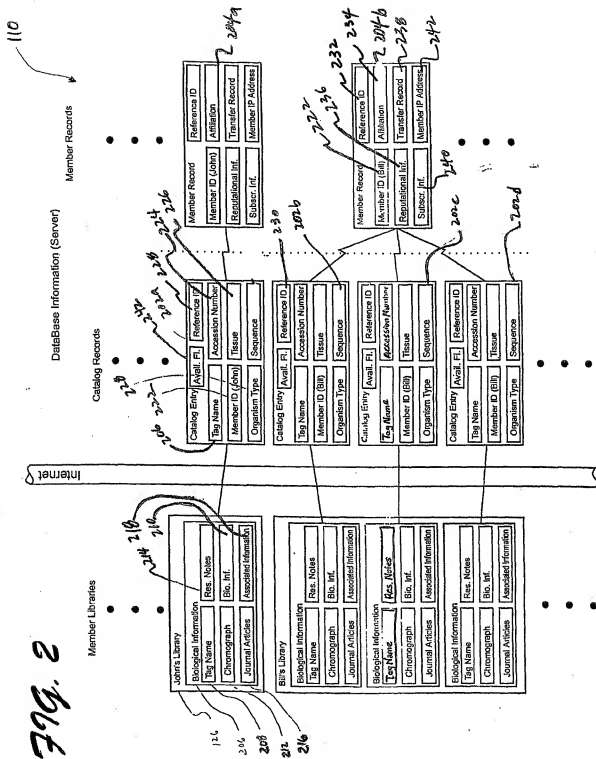
when it is determined that the network device on which the desired biological information is stored is communicably connected and that the searching member is not communicably connected to the network host, transmitting to the network device on which the desired biological information is stored the stored request to transfer a copy of the desired biological information to the network host;

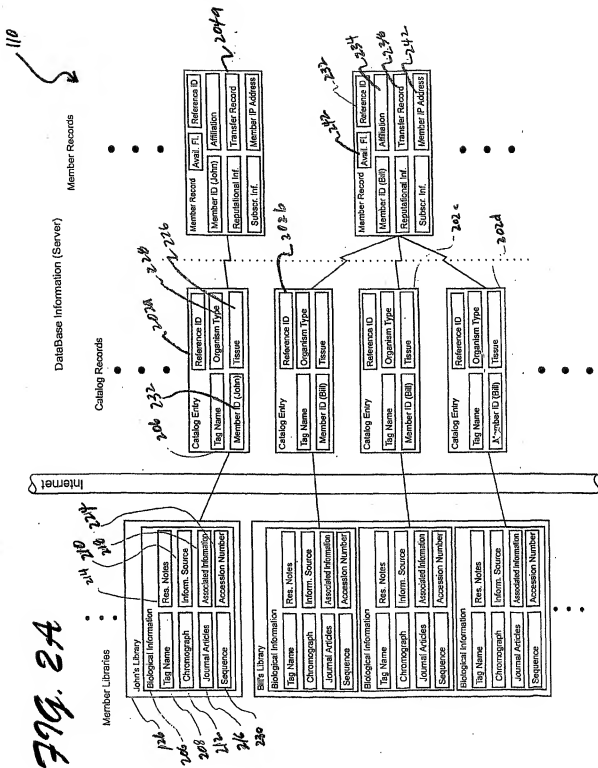
storing the desired biological information on the network host; and

transmitting the desired biological information from the network host to the searching member when the searching member is next communicably connected to the network host.

Fig. 1







302 304 306 308 300

Chat Library Search Transfer

Search by sequence string  310

Gene name  312

User code

Accession number  314

Organism  316

Tissue  318

**FIND IT**

Search Results

Sequence tag name	% match	Gene	User	Organism	Tissue	ACC. #
DXB11111	100	p53	Dave1	Xenopus		n/a
p53rac1h	97	p53	Chet1	Xenopus	Brain	A123324253
grabdr	85	p53	J234i	Drosophila	Liver	A123324553

Figure 3



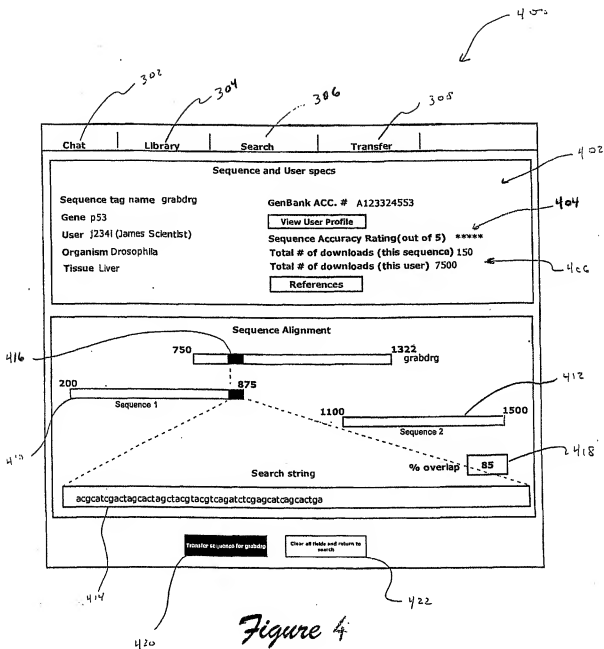


Figure 4

FIG. 5A

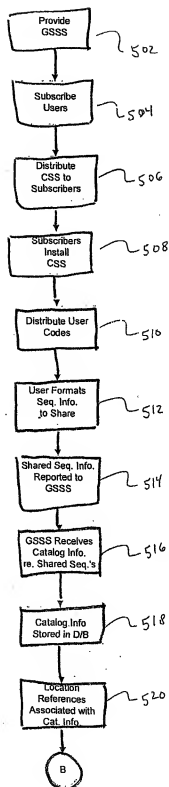


Fig. 5 B

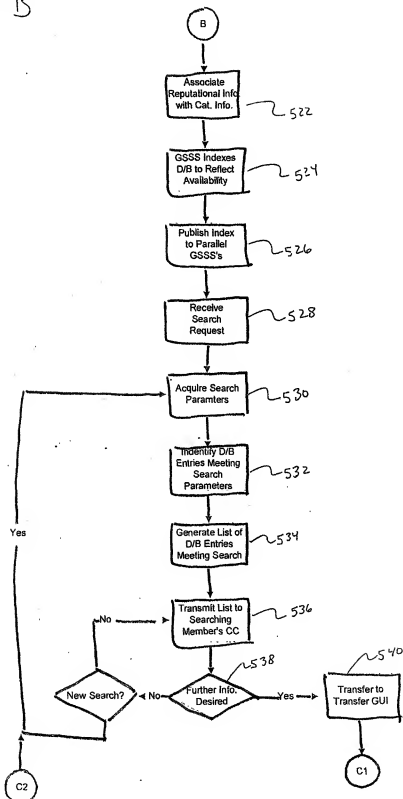
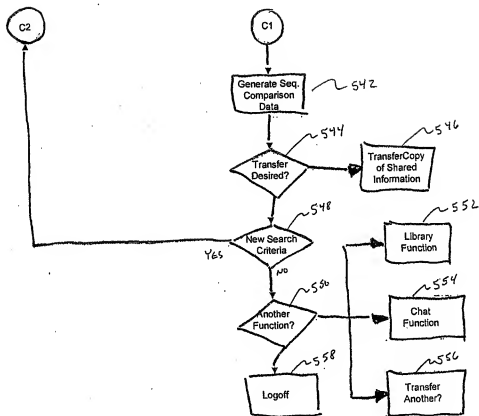


FIG. 5C



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/25056

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G06F 17/80

US CL : 705/8

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 705/8

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EAST (US Patents, Derwent, EPO, JPO), PROQUEST (all databases)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,970,500 A (SABATINI et al) 19 October 1999, col. 10, line 59 - col.11, line 7, and lines 35-46.	1-52
Y,P	US 6,125,383 A (GLYNIAS et al) 26 September 2000, see abstract, col. 3, lines 48-58, col. 6, lines 37-57, col. 11, lines 1-6 and col. 18, lines 44-55	1-52
A,E	US 6,308,170 B1 (BALABAN) 23 October 2001, see abstract.	1-52
A	US 5,966,712 A (SABATINI et al) 12 October 1999, see abstract.	1-52



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents	*T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A <sup>1</sup> document defining the general state of the art which is not considered to be of particular relevance	*X	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*Z <sup>1</sup> earlier document published on or after the international filing date	*Y	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L <sup>1</sup> document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)		
*O <sup>1</sup> document referring to an oral disclosure, use, exhibition or other means		
*P <sup>1</sup> document published prior to the international filing date but later than the priority date claimed	*A <sup>2</sup>	document number of the same patent family

Date of the actual completion of the international search

12 DECEMBER 2001

Date of mailing of the international search report

16 JAN 2002

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 505-5250

Authorized officer

TARIQ HAFIZ

Telephone No. (703) 505-7045

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US01/25956

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KUSKA, BOB. Online databases bring research to the desktop. Journal of the National Cancer Institute. 19 April 2000. Vol. 92. No. 8. pages 600-601.	1-52
A	PAPANIKOLAW, JIM. Bioinformatics emerges as key technology for developing new drugs. Chemical Market Reporter. 24 May 1999. Vol. 255. No. 21. page 22.	1-52